软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(3):495−497 [doi: 10.13328/j.cnki.jos.005701]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

智能数据管理与分析技术专刊前言*

樊文飞[1,2], 王国仁[3], 王朝坤[4]

[1](School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, UK)

[2](北京航天航空大学 计算机学院,北京 100191)

[3](北京理工大学 计算机学院,北京 100081)

[4](清华大学 软件学院,北京 100084)

通讯作者: 王朝坤, E-mail: chaokun@tsinghua.edu.cn

中文引用格式: 樊文飞,王国仁,王朝坤.智能数据管理与分析技术专刊前言.软件学报,2019,30(3):495−497. http://www.jos.org.cn/1000-9825/5701.htm

数据管理与智能计算的深度融合已经成为大数据时代顺利前行的迫切需求.智能数据管理旨在“为数据增添智能”,是数据科学与技术的重要基石,更是大数据产业蓬勃发展的关键支撑.一方面,将新一代人工智能方法应用于先进数据管理技术,尝试探索和突破智能数据管理与分析的理论体系、技术方法及系统平台,已经成为数据管理领域的新兴研究方向;另一方面,研发面向人工智能的数据库基础软件,为新一代人工智能技术的研发和广泛应用提供海量数据的有效存储、查询、分析和挖掘等的系统支持,亦是国家科技创新的决定性因素.智能数据管理与分析领域日益得到学术界和工业界的普遍关注,其理论、技术和方法亟待深入地探索与思考.目前,针对智能数据管理与分析的研究仍然处于起步阶段,有很多需要研究的问题.

本专刊公开征文,共收到投稿 38 篇(包括第 35 届中国数据库学术会议(NDBC 2018)推荐的 12 篇高质量论文).其中,37 篇论文通过了形式审查,内容涉及智能数据管理与分析技术和应用.特约编辑先后邀请了 70 多位专家参与审稿工作,每篇投稿至少邀请 2 位专家进行评审.稿件经初审、复审、NDBC 2018 会议宣读和终审 4 个阶段,历时 5 个月,最终有 20 篇论文入选本专刊.根据主题,这些论文可以分为 4 组.

(1) 智能图数据管理技术.

《大规模 RDF 图数据上高效率分布式查询处理》提出基于 MapReduce 框架的查询处理器 SDec 有效回答大规模 RDF 智能图数据上的 SPARQL 基本图模式查询.

《基于规则的最短路径查询算法》设计了一种基于最优子路径的前向扩展算法,可快速求解基于规则的最短路径查询问题,并进一步设计了基于最短优先策略的前向扩展算法.

《基于角色发现的动态信息网络结构演化分析》使用角色来量化动态网络的结构,并给出两种解释角色的方法;将动态网络结构预测问题转换为角色预测问题,提出基于潜在角色的动态网络结构预测方法.

《复杂条件下的社区搜索方法》给出了条件社区搜索问题的形式化定义,使用布尔表达式表示搜索条件;进而提出解决条件社区搜索问题的通用框架及其优化方法,将条件社区搜索分解为多个单项条件社区搜索.

《基于事件的社交网络上的双边偏好稳态规划》研究了如何为社交网络中的用户规划感兴趣的事件,提出了双边偏好稳态规划算法,考虑了用户和事件彼此间的偏好效用.

(2) 智能数据管理方法与工具.

《基于时效规则的数据修复方法》针对同一实体对应的多条记录存在时间戳缺失或不精确条件下的数据时效修复问题,给出了通用的状态类型时效规则提取算法,以及基于时效规则的数据时效修复算法.

《劣质数据上代价敏感决策树的建立》定义了劣质数据上代价敏感决策树的建立问题,提出了 3 种融合数据清洗算法的代价敏感决策树建立方法.

收稿时间: 2018-12-19

《两两比较模型的 Why-not 问题解释及排序》从利用两两比较方法寻找函数依赖的算法中得到启发,将两两比较方法、统计学方法以及机器学习方法进行结合,针对 Why-not 问题寻找解释并对解释进行排序.

《差分隐私的数据流关键模式挖掘方法》提出了一种满足差分隐私的数据流关键模式挖掘算法,既考虑了隐私和数据效用之间的权衡,又考虑了挖掘时间和维护开销之间的权衡.

《基于网格耦合的数据流聚类》针对现有数据流聚类算法在实时处理高速、大量的数据流时聚类效率和精度不高的问题,提出了一种基于网格耦合和核心网格的数据流聚类算法.

《分布式异构数据库数据同步工具》提出了一种基于 MySQL 二进制日志还原 SQL 的方法,设计了日志解析器和日志还原器,可针对不同事件进行日志解析,并依据相应规则还原生成可执行的 SQL 语句.

(3) 智能数据分析技术及应用.

《面向通用模型的高可用性步态周期分析方法》提出了一种结合波峰波谷检测与阈值空间的高可用性步态周期分析方法,通过自动求解预估值,并构建自适应区间,根据通用步态模型对缺乏上述信息的未知步态数据进行切分与分析,能更便利准确地求解步态周期数据.

《CNN 多位置穿戴式传感器人体活动识别》针对现有二维卷积输入构建方法中对多位置三轴向传感器相同轴向数据之间的空间依赖性挖掘不足的现象,提出了多层卷积神经网络模型并应用于基于传感器数据的人体活动识别.

《改进的 SSD 航拍目标检测方法》针对无人机场景下目标分辨率低、尺度变化大等问题,在 SSD 目标检测算法的基础上,采用表征能力更强的残差网络进行基准网络替换,引入跳跃连接机制降低提取特征的冗余度,引入不同分类层的特征信息融合机制来有机结合网络结构中低层视觉特征与高层语义特征.

《面向交通流量预测的多组件时空图卷积网络》提出了一种多组件时空图卷积网络,该模型结合图卷积和标准卷积构造时空卷积块来同时捕获交通数据的时空特性.

《时空依赖的城市道路旅行时间预测》针对传统旅行时间预测模型难以引入多源特征的问题,提出了两阶段的旅行时间预测框架,有效提取路段间上下游依赖关系,且整合了天气日期等外部特征.

(4) 智能数据分析方法与进展.

《面向高维特征和多分类的分布式梯度提升树》证明了特征并行策略更适合高维和多分类场景，提出了一种使用特征并行的分布式梯度提升树算法.

《因子分解机模型研究综述》从准确性和性能两个方面总结了因子分解机模型存在的基本问题和近年来的研究进展,综述了适用于因子分解机模型求解的 4 种代表性优化算法.

《因子分解机模型的宽度和深度扩展研究》从特征的高阶交互、场交互、层次交互与传统模型的集成学习,以及特征工程角度讨论了 FM 模型的宽度扩展,从与深度学习模型等集成的角度,详细阐述了 FM 模型的深度扩展,同时概括比较了 FM 模型的优化学习方法和基于不同并行与分布式计算框架的实现.

《基于强化学习的金融交易系统研究与发展》以金融领域常用的强化学习模型发展为脉络,对交易系统、自适应算法、交易策略等方面诸多研究成果进行了综述.

本专刊主要面向数据库、数据挖掘、大数据、机器学习等多领域的研究人员和工程人员,反映了我国学者在智能数据管理与技术领域最新的研究进展.感谢《软件学报》编委会和数据库专委会对专刊工作的指导和帮助,感谢专刊全体评审专家及时、耐心、细致的评审工作,感谢踊跃投稿的所有作者.希望本专刊能够对智能数据管理与分析相关领域的研究工作有所促进.

樊文飞(1963－),男,博士,教授,博士生导师,英国皇家学会院士(FRS),欧洲科学院院士(MAE),爱丁堡皇家学会院士(FRSE),美国计算机协会会士(ACM Fellow).现任英国爱丁堡大学信息学院首席教授.曾获得英国皇家学会 Wolfson 研究成果奖(2018)、欧洲研究委员会 ERC Advanced Fellowship(2015)、英国 Roger Needham 奖(2008)、海外杰出青年学者(2003)、美国 CAREER Award(2001),Elsevier 网络科学刊物年度最佳论文和最杰出作者奖(2002),SIGMOD 2017 突出研究奖以及数据管理四大国际顶级理论与系统会议的时间检验奖和最佳论文奖: Alberto O. Mendelzon 时间检验奖/ACM PODS 十年最佳论文奖(2010 和 2015),ACM SIGMOD(2017)、VLDB(2010)和 ICDE(2007)最佳论文奖.目前主要研究领域为数据库理论与系统,包括大数据,数据质量,数据集成,分布式计算,查询语言,推荐系统,社会网络分析与精准营销.

王国仁(1966－),男,博士,教授,博士生导师,分别于 1988 年、1991 年和 1996 年获得东北大学计算机专业学士、硕士和博士学位.主持国家杰出青年科学基金、国家自然科学基金重点项目和广东联合基金重点项目、国家 863 计划项目等 20 余项,发表学术论文 100 余篇,主要研究领域为不确定数据管理,数据密集型计算,可视媒体数据管理与分析,非结构化数据管理,分布式查询处理与优化技术(主要包括传感器网络和 P2P 对等计算),生物信息学等.

王朝坤(1976－),男,博士,长聘副教授,博士生导师,北京电子学会理事,分别于 1997 年、2000 年和 2005 年获得哈尔滨工业大学计算机科学与工程专业学士、计算数学专业硕士和计算机软件与理论专业博士学位.作为主持人正在承担国家自然科学基金、国家重点研发计划课题等 4 项科研项目,发表论文 100 余篇,获得最佳论文奖 6 项,授权中国发明专利 20 项.获国家科技进步二等奖 1 项,省部级科研奖励 3 项.主要研究领域为社交网络及图数据管理,非结构化大数据技术与系统.